Les 4. Hypothesetoetsen

Onderzoekstechnieken

Jens Buysse   Wim De Bruyn   Wim Goedertier   Bert Van Vreckem

AJ 2018-2019

HO
GENT

What's on the menu today?

Toetsen van hypothesen

Werkwijze

Overschrijdingskans

Kritieke gebied

Voorbeelden

De t-toets

Fouten in hypothesetoetsen

Effectgrootte

HO GENT

Toetsen van hypothesen

HO
GENT

De statistische test voor een hypothese

Hypothese Idee waarvan nog bewezen moet worden dat het juist is: uitspraak over numerieke waarde van een populatieparameter

Hypothesetest controle van een uitspraak over de waarden van één of meerdere populatieparameters

Nulhypothese (H_0) Deze hypothese proberen we te ontkrachten door redenering in het ongerijmde

Alternatieve hypothese (H_1, H_a) Deze hypothese willen we aantonen

HO GENT

Elementen bij toetsingsprocedure

Toetsingsgrootheid De variabele die berekend wordt uit de steekproef (ook: teststatistiek)

Aanvaardingsgebied Het gebied van waarden die de nulhypothese ondersteunt

Kritieke of Verwerpingsgebied Het gebied van waarden die de nulhypothese verwerpt

Significantieniveau De maximale kans dat H_0 onterecht verworpen wordt

HO GENT

Werkwijze

Werkwijze

1. Bepalen van de hypothesen (H_0 en H_1)
2. Vastlegen significantieniveau (α)
3. Toetsingsgrootheid berekenen
4. Het kritieke gebied of de overschrijdingskans bepalen
5. Conclusies trekken

HO
GENT

hypothesen over superhelden

Een superheld redt 3.3 mensen per dag

Bron: `http://www.cracked.com/quick-fixes/4-people-who-saved-day-while-dressed-as-superheroes/`

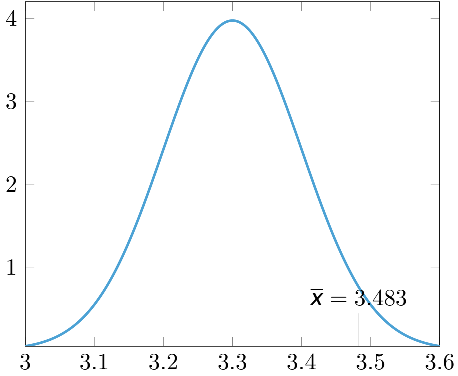
Stel, in een periode van 30 dagen werden er gemiddeld $3,483$ mensen per dag gered ($\overline{x} = 3,483$, $n = 30$)

1. Hypothese: $H_0 : \mu = 3,3$; $H_1 : \mu > 3,3$
2. Significantieniveau: $\alpha = 0,05$
3. Steekproefgrootheid: $\overline{x} = 3,483$

Populatiestandaardafwijking (verondersteld gekend): $\sigma = 0,55$

Berekenen toetsingsgrootheid

Uit centrale limietstelling volgt: $M \sim Nor(\mu = 3,3; \sigma = \frac{0,55}{\sqrt{30}} = 0,1)$



Overschrijdingskans

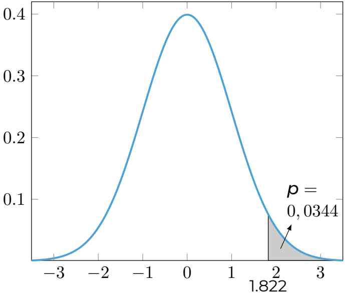
Overschrijdingskans

De p-waarde is de kans, indien de nulhypothese waar is, om een waarde te verkrijgen voor de toetsingsgrootheid die minstens even extreem is als de geobserveerde waarde.

- p-waarde $< \alpha \Rightarrow H_0$ verwerpen: de gevonden waarde voor $\overline{x}$ is te extreem;
- p-waarde $\geq \alpha \Rightarrow H_0$ niet verwerpen: de gevonden waarde voor $\overline{x}$ kan nog verklaard worden door toeval.

HO GENT

Overschrijdingskans



$$P(M > 3,483) = P\left(Z > \frac{3,483 - 3,3}{\frac{\sigma}{\sqrt{n}}}\right) = P(Z > 1,822) = 0,0344$$

HO GENT

Kritieke gebied

Kritieke gebied

Het kritieke gebied is de verzameling van alle waarden voor de toetsingsgrootheid waarbij de nulhypothese kan verworpen worden.

We zoeken een grenswaarde g waarvoor geldt:

$$P(M > g) = \alpha$$

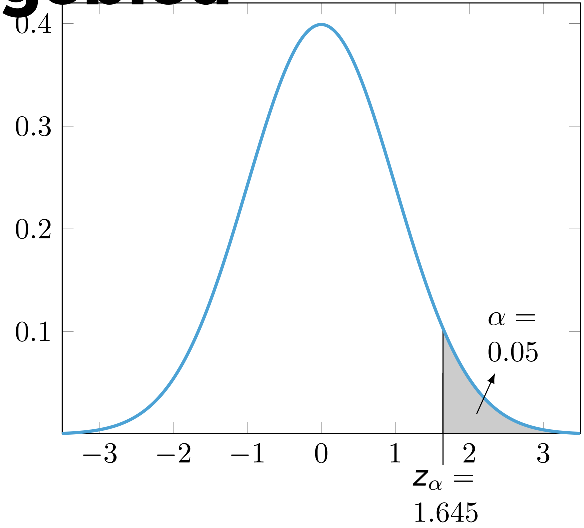
Bepaal z_α waarvoor geldt:

$$P(Z > z_\alpha) = \alpha \Rightarrow g = \mu + z_\alpha . \frac{\sigma}{\sqrt{n}}$$

- Links van g: aanvaardingsgebied (H_0 niet verwerpen)
- Rechts van g: kritieke gebied (H_0 wel verwerpen)

HO GENT

Kritieke gebied



HO GENT

significantieniveau $\alpha = 0.05 \Rightarrow$ grenswaarde z_α=1.645

(z_α =`qnorm(0.95)`)

Linkszijdig toetsen

Wat zou je in de verg. moeten veranderen opdat je de correcte kritieke waarde zou berekenen?

Linkszijdig toetsen

Wat zou je in de verg. moeten veranderen opdat je de correcte kritieke waarde zou berekenen? Antwoord:

$$g = \mu - z \times \frac{\sigma}{\sqrt{n}}$$

want

$$P(M < g) = P\left(Z < \frac{g - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0.05$$

HO GENT

Linkszijdig toetsen

Wegens de symmetrieregel kunnen we zeggen

$$P\left(Z > -\left(\frac{g-\mu}{\frac{\sigma}{\sqrt{n}}}\right)\right) = 0.05$$

De z-waarde die ermee overeen komt is 1.645 dus hebben we

$$z = \frac{-g+\mu}{\frac{\sigma}{\sqrt{n}}} \Leftrightarrow -g = \frac{\sigma}{\sqrt{n}}z - \mu$$

$$\Leftrightarrow g = \mu - z\frac{\sigma}{\sqrt{n}}$$

Tweezijdig toetsen

Soms kan het ook zijn dat er tweezijdig moet getoetst worden. Er moeten dan twee kritieke grenswaarden berekend worden namelijk de linker- en de rechter grenswaarden.

$$g = \mu \pm z \times \frac{\sigma}{\sqrt{n}} \tag{1}$$

HO GENT

Samenvatting

Doel	Test op gemiddelde waarde μ van de populatie aan de hand van een steekproef van n onafhankelijke steekproefwaarden		
Voorwaarde	De populatie is willekeurig verdeeld, n voldoende groot		
Type test	Tweezijdig	Eenzijdig links	Eenzijdig rechts
H_0	$\mu = \mu_0$	$\mu = \mu_0$	$\mu = \mu_0$
H_1	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$
Verwerpingsgebied	$\lvert z \rvert > g$	$z < -g$	$z > g$
Teststatistiek		$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	

Tabel: Samenvatting mogelijke toetsen

Voorbeelden

Voorbeeld 1

Bij een aselecte steekproef van 50 waarnemingen vinden we volgende grootheden: gemiddelde $\overline{x} = 25$ en standaardafwijking s = $\sqrt{55} = 7.41$ We willen weten of er reden is om aan te nemen dat gemiddelde van de populatie kleiner is dan 27.

HO
GENT

Voorbeeld 1

1. Bepalen van de hypothesen
   $H_0 : \mu = 27$ en $H_1 : \mu < 27$.
2. Vastleggen significantieniveau
   $\alpha = 0.05$ en $n = 50$
3. Toetsingsgrootheiden & waarde: steekproefgemiddelde
   $\overline{x} = 25$

HO
GENT

Voorbeeld 1 (vervolg)

4a Overschrijdingskans

Volgens de centrale limietstelling geldt:

$$M \sim Nor(\mu = 27, \frac{\sigma}{\sqrt{n}})$$

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{25 - 27}{\sqrt{\frac{55}{50}}} \approx -1.91$$

We vinden dus een overschrijdingskans van het gemiddelde 0.0281.

Bij een significantieniveau van 0.05 mogen we H_0 verwerpen.

Voorbeeld 1 (vervolg)

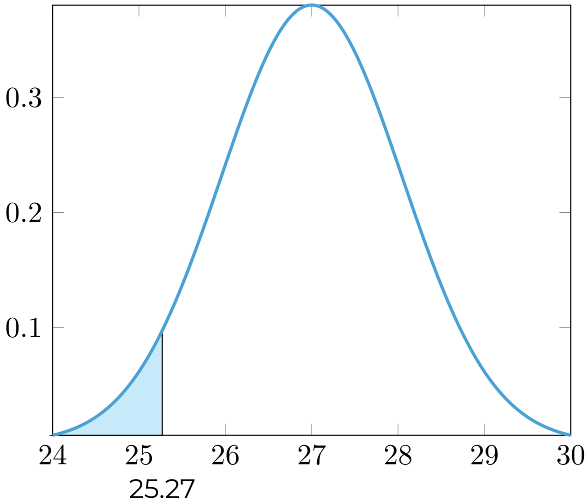
4b Bereken en teken kritiek gebied

$$
\begin{aligned}
g &= \mu - z \times \frac{\sigma}{\sqrt{n}} \\
&= 27 - 1.645 \times \sqrt{\frac{\sigma}{n}} \\
&= 25.27470944
\end{aligned}
$$

We vinden dus dat $\overline{x} < g$ en dus kunnen we H_0 verwerpen.

HO GENT

Voorbeeld 1 (vervolg)



HO
GENT

Voorbeeld 1 (vervolg)

5 Conclusie
We besluiten op basis van de steekproef dat $\mu < 27$ met een significantieniveau van $0,05$

HO GENT

Voorbeeld 2

In een onderzoek naar het kleingeld dat in de zakken van van onze superhelden zit, wordt er van uit gegaan dat ze gemiddeld €25 op zak hebben. We gaan ervan uit dat we een gekende spreiding van $\sigma = 7$ hebben. Verder zijn de gegevens van de aselecte steekproef van omvang $n = 64$ beschikbaar met gemiddeld zakgeld $\overline{x}$ van €23. Neem als significantieniveau $\alpha = 0.05$.

Voorbeeld 2

1. Bepalen van de hypothesen
   $H_0 : \mu = 25$ en $H_1 : \mu \neq 25$
2. Vastleggen significantieniveau
   $\alpha = 0.05$ en $n = 64$.
3. Toetsingsgrootheden & waarde: $\bar{x} = 23$

HO GENT

Voorbeeld 2 (vervolg)

4b Bereken kritieke gebied

$$g_1 = \mu - z \times \frac{\sigma}{\sqrt{n}} = 23.28$$

$$g_2 = \mu + z \times \frac{\sigma}{\sqrt{n}} = 26.72$$

We vinden dat $\overline{x}$ in het kritieke gebied ligt (want $23 < 23.28$) dus mogen we H_0 verwerpen.

5 We kunnen op basis van deze steekproef besluiten dat de superhelden *minder* dan 25 € op zak hebben, met een significantieniveau van 5%

HO GENT

De *t*-toets

Veronderstellingen z-toets

- De steekproef moet aselect zijn
- De steekproef moet voldoende groot zijn ($n \geq 30$)
- De variatie van de toetsingsgrootheid moet normaal verdeeld zijn
- We veronderstellen dat de standaardafwijking van de populatie, σ, gekend is

Als deze veronderstellingen niet gelden, mag je de z-toets niet gebruiken!

De t-toets

Bepalen kritieke grenswaarde:

$$g = \mu \pm t \times \frac{s}{\sqrt{n}}$$

- *t*-waarde wordt afgeleid uit de Student-*t* verdeling, hangt af van aantal *vrijheidsgraden*, $n - 1$
- Op te zoeken in *t*-tabel of met R-functie `pt`
- Procedure hypothesetoets is verder identiek aan *z*-toets

HO
GENT

Vergelijken van twee steekproeven

Is steekproefgemiddelde van twee steekproeven significant verschillend?

- Onafhankelijke steekproeven
- Gepaarde steekproeven

Onafhankelijke steekproeven

In een klinisch onderzoek wil men nagaan of een nieuw medicijn als bijwerking een verminderde reactiesnelheid heeft.

- Controlegroep: 6 deelnemers krijgen placebo
- Interventiegroep: 6 deelnemers krijgen medicijn

Vervolgens wordt reactiesnelheid gemeten:

- Controlegroep: 91, 87, 99, 77, 88, 91 ($\bar{x} = 88,83$)
- Interventiegroep: 101, 110, 103, 93, 99, 104 ($\bar{y} = 101,67$)

Zijn er significante verschillen tussen de interventie- en controlegroep?

HO GENT

Onafhankelijke steekproeven

1. Hypothese:
   - H_0: $\mu_1 - \mu_2 = 0$
   - H_1: $\mu_1 - \mu_2 < 0$
2. Significantieniveau: $\alpha = 0,05$
3. Steekproefgrootheid:
   - $\bar{x} - \bar{y} = -12,833$
   - $\bar{x} =$ schatting voor μ_1 (controlegroep)
   - $\bar{y} =$ schatting voor μ_2 (interventiegroep)

In R:

```
> controle <-  c(91, 87, 99, 77, 88, 91)
> interventie <- c(101, 110, 103, 93, 99, 104)
> t.test(controle, interventie, alternative="less",
                                      mu=0, conf.level=0.95)
```


HO GENT

Onafhankelijke steekproeven

```
Welch Two Sample t-test

data:  controle and interventie
t = -3.4456, df = 9.4797, p-value = 0.003391
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf -6.044949
sample estimates:
mean of x mean of y
88.83333 101.66667
```

OPGELET! Dit heeeft niets te maken met aanvaardingsgebied of kritiek gebied

$\overline{x} - \overline{y} = -12,833$ komt overeen met $t = -3,4456$

$df = 9,48$ wordt berekend door `t.test()` op basis van x en y

$p = 0,003391 < \alpha = 0,05$ dus H_0 verworpen (significant verschil)

HO GENT

Gepaarde steekproef

In een studie werd nagegaan of auto's die rijden op benzine met additieven ook een lager verbruik hebben.
Bij 10 auto's werd het verbruik gemeten (uitgedrukt in mijl per gallon) voor beide soorten benzine:

Auto	1	2	3	4	5	6	7	8	9	10
Gewone benzine	16	20	21	22	23	22	27	25	27	28
Met additieven	19	22	24	24	25	25	25	26	28	32

Gepaarde steekproef

1. Hypothese:
   - H_0: $\overline{x - y} = 0$
   - H_1: $\overline{x - y} > 0$
2. Significantieniveau: $\alpha = 0,05$
3. Steekproefgrootheid:
   - $\overline{x - y}$
   - $x =$ mijl per gallon met additieven ($\overline{x} = 25,1$)
   - $y =$ mijl per gallon met gewone benzine ($\overline{y} = 23,1$)

In R:

```
> gewone     <- c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> additieven <-c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> t.test(additieven, gewone, alternative="greater",
                                   paired=TRUE, mu=0, conf.level=0.95)
```

Gepaarde steekproef

```
Paired t-test

data:  additieven and gewone
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true difference in means is greater than 0
```

~~95 percent confidence interval:~~

~~1.180207 Inf~~

OPGELET! Dit heeeft niets te maken met aanvaardingsgebied of kritiek gebied

```
sample estimates:
mean of the differences
2
```

$\bar{x} - \bar{y} = 2$ komt overeen met $t = 4,4721$

$p = 0,0007749 < \alpha = 0,05$ dus H_0 verworpen (significant verschil)

Fouten in hypothesetoetsen

HO
GENT

Fouten in hypothesetoetsen

	Realiteit (onbekend)	
Conclusie toets	H_0 **correct**	H_1 **correct**
H_0 **geaccepteerd**	Juist	Fout van type II
H_0 **verworpen**	Fout van type I	Juist

P(type I error) = α (= significantieniveau)
P(type II error) = β
β berekenen is ***niet*** triviaal, maar als $\alpha \searrow$ dan $\beta \nearrow$

HO
GENT

Effectgrootte

Effectgrootte

De effectgrootte is een metriek die uitdrukt hoe groot het verschil tussen twee groepen is

- Controlegroep vs. interventiegroep
- Kan je gebruiken naast hypothesetoets
- Vaak gebruikt in onderwijswetenschappen
- Er zijn verschillende definities, hier: *Cohen's d*

Cohen's *d*

$$d = \frac{\overline{x_1} - \overline{x_2}}{s}$$

met $\overline{x_1}, \overline{x_2}$ gemiddelde van de steekproeven
en s een gecombineerde standaardafwijking over beide groepen:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

met n_1, n_2 de steekproefgroottes, s_1, s_2 de standaardafwijking van beide groepen

HO GENT

Interpretatie Cohen's *d*

$\lvert d \rvert$	**Effectgrootte**
0,01	zeer klein
0,2	klein
0,5	matig
0,8	groot
1,2	zeer groot
2,0	enorm

In onderwijswetenschappen (John Hattie):

- 0,4 = kantelpunt voor gewenste effecten
- effectgrootte $d = 1$: leerstof van 1j op 6m verwerken!

Zie bvb. `http://www.evidencebasedteaching.org.au/hatties-2017-updated-list/`

HO GENT

Typisch opzet onderzoek in onderwijs

- Onderzoeksvraag: Is X een goede leerstrategie, m.a.w. heeft dit een positief effect op eindresultaten?
- Controlegroep gebruikt “normale”, “klassieke” techniek
- In de interventiegroep wordt X toegepast
- Achteraf volgt evaluatiemoment
- Scores bepalen, *d* berekenen